Broadfoot, P. (2017). Time to tame the leviathan? Perspectives and possibilities for a new era in assessment. *Assessment in Education: Principles, Policy and Practice*, *24*(3), 415-423. https://doi.org/10.1080/0969594X.2017.1336428

Time to tame the leviathan? Perspectives and possibilities for a new era in assessment.

Patricia Broadfoot
University of Bristol

Introduction

The focal paper in this special issue argues that there are few issues more important in education than the relationship between assessment and learning. On the one hand, learning is *the* ultimate goal of all educational interventions; on the other, assessment has become a leviathan that constrains and shapes virtually all of that formal educational activity.[1] Thus the argument in '*Assessment and Learning: fields apart*' that theories of learning and theories of assessment should be developing more closely with each other is very welcome. Indeed it is arguably overdue. From an objective viewpoint, it seems extraordinary that the overwhelming influence of various kinds of assessment on contemporary educational systems and practices has not been subject to more rigorous scrutiny in terms of the extensive research that exists on how best to encourage learning. As Baird et al assert in their focal paper, '*educational assessments affect what and how learning occurs'.* One of the reasons for this is that assessment data provide such a convenient proxy for presumed quality in education. The result is that all too often its use has almost nothing to do with learning and everything to do with power and control. This is a situation that is intensifying with every year that goes by as globalisation steadily raises the educational stakes in relation to individual, institutional and national achievements. It is therefore a challenge that is becoming increasingly pressing for the professional educational community. I am therefore delighted that the editors of this Special Issue of *Assessment in Education* have thrown down a collective gauntlet for scholars to address the issue with some urgency.

It is clearly high time that educational research – a field that according to Baird et al has yet to reach maturity – should make a sustained effort to grasp this particular nettle. But, as the authors make very clear in their excellent focal paper, it is indeed a nettle, since the effort to tease out the relationship between these two central pillars of education is fraught with difficulty. As they demonstrate clearly in their article, such an effort must engage with the competing paradigms of assessment and psychometrics. It must embrace the implications of the tensions between various theories of learning and between the different purposes of assessment. It must also address the

[1]The Oxford English dictionary defines leviathan as 'A thing that is very large or powerful'.

generic tensions between quantitative and statistical *measurement* data with its associated assumptions that performance can be measured in the same way as a physical entity, and more descriptive, qualitative approaches to assessment that make fewer any assumptions about underlying constructs but are intended rather to provide more holistic forms of reporting. Last but not least, it must address the inevitable tension between different priorities in the construction of assessments depending on the context in which the data are to be used and the cultural traditions that shape the conceptualising of assessment priorities and criteria of quality. In view of all these theoretical and practical differences that surround the business of assessing learning, it is perhaps not as surprising, as the authors suggest, that relatively little progress seems to have been made on this agenda and that two of the central aspects of the educational enterprise are currently 'fields apart'.

In responding to this article it seems to me there are three central issues that merit further discussion. These are:

- which are the really key issues in this relationship that need urgent attention as opposed to those issues which are more 'academic niceties' and why?
- is there more to be said about learning theory that is relevant to assessment than is discussed in the stimulus article?
- are there novel theoretical or practical developments – either in the field of education or more generally – that might facilitate a more fruitful coming together of the two fields of assessment and learning?

Which issues in the relationship between assessment and learning are particularly urgent?

By common consent, a great deal of educational assessment is now 'high stakes'. In their discussion of international large-scale assessments (ILSA), Baird et al. ably describe the powerful influence of such tests on national educational policy, especially national curricula. They argue that whilst the impact of such tests is less direct than that of individual summative assessments, they are nevertheless extremely powerful in that the data produced are given a degree of weight that the inevitable technical limitations associated with such comparisons would argue against. In addition to the practical difficulties of achieving any comparison of like with like, they also point out the more profound assessment issues associated with the psychometric paradigm such as the difficulty of simultaneously achieving both invariance and validity in an attempt to quantify educational attainment. Whilst it might appear possible to make some sort of international comparison of

performance, any attempt to extrapolate from such observables to the underlying constructs being measured is therefore fraught with difficulty.

It is perhaps for this reason that, as the authors suggest, surprisingly little attempt has so far been made to use the test data generated in order to understand more about learning per se. Instead, they argue, the "use of educational assessment scores in the many second-order systems that have been generated for accountability purposes has created a gulf between what is assessed and how the data are used*";* that "use of assessment data only as signifiers is detrimental to this agenda" (i.e. the educational purpose of assessment). Indeed they offer a devastating critique of the utility of such tests, especially in the international context and when designed to be unrelated to particular curriculum content. Apart from the practical problems they mention of making data-collection comparable between countries, the melange of theoretical, empirical, expert and political inputs into the test construction process almost certainly ensures that the results have little relationship to any generalisable construct. It is certainly the case that this failure to link the design and subsequent data-interpretation of such tests to any relevant theories of learning, lends great weight to the authors' central argument that there needs to be a much stronger correspondence between learning and assessment theories if assessment is to maximise its contribution to the educational project in its broadest sense.

Governments and similar bodies appear to have few scruples concerning the use of such national and international test data despite knowing little about their potential impact on the core business of education – student learning. Perceived legitimacy and expediency appear to rule the day. To put it crudely, if it 'works' for the intended purpose, that is good enough. The onus is therefore on the educational research community to demonstrate the opportunity cost of *not* paying more attention to the implications that any particular assessment has for helping or hindering learning.

Interestingly, the authors' complement their analysis of the problems besetting the relationship between international tests and learning theory with a discussion of the very different rationale and practice of Assessment for Learning (AfL). Here surely, we will find better support for the paper's central argument "that if assessments are to serve the goals of education, then theories of learning and assessment should be developing more closely with each other" (p. 1). For if a clear goal for all assessment specialists, whether they be commercial testing organisations at one end of the spectrum or classroom teachers at the other, is to create 'tests worth teaching' then it follows that assumptions about learning must be built in to such projects. The rapid international spread of the AfL paradigm is testimony to its perceived value in practice, its positive potential to support and encourage

learning. However, as the authors suggests, not only is their disagreement as to the scale of such potential, more significantly "the theoretical underpinnings are often left implicit so that AfL becomes a series of classroom practices (for example, learning intentions, wait time, comment-only marking) justified and validated in terms of 'what works'" (p.37) rather than any particular learning theory, leading to their conclusion that "the theoretical basis for AfL is eclectic and often implicit".

Of the many important challenges for the assessment community that are addressed in this paper, it would therefore seem that the really key issue in the relationship between assessment and learning theory that needs urgent attention is to understand better how to create 'tests worth teaching' in every aspect of assessment activity be these individual summative tests, formal examinations or national and international surveys of achievement. As Baird et al assert, "educational purposes should be the priority" (p. 31). The implication is clear – that there is an urgent need to focus more effort on the analysis of different kinds of assessment data with a view to understanding more about what they reveal about patterns and processes in learning.

Is there more to be said about learning theory that is relevant to assessment?

In Table 1, Baird et al list four theoretical orientations that can be linked to different approaches to formative assessment – behaviourist; cognitive constructivist; social constructivist and socio-cultural, listing also some of the leading assessment scholars in these various theoretical camps. It is important to say firstly, how welcome the recent emergence of the more post-modern approaches listed here are, offering as they do the opportunity for philosophical debates about the nature of individual reality which are far more fundamental than the more well-established behaviourist and cognitive traditions in assessment. Given the powerful critiques that the authors offer in their focal paper concerning the shortcomings of traditional psychometric approaches, this discussion opens new horizons in the search for a more sophisticated relationship between assessment and learning perhaps – as the authors suggest – for an evolution in the discipline of assessment that mirrors the way in which natural science evolves, and heralding a much more robust rapprochement between assessment and its fundamental purpose of enhancing learning.

Important as these theoretical perspectives are however, they are far from the whole story concerning the relationship between assessment and learning. Other perspectives, such as those discussed in an early review of the research literature on this subject in a now classic paper by Crooks (1988) offer a different lens with which to approach the relationship between assessment

and learning. Crooks' review highlights the importance on learner dispositions and emotional state as key drivers – the affective domain. Assessment, he suggests

> guides their [the learners'] judgment of what is important to learn, affects their motivation and self-perceptions of competence, structures their approaches to and timing of personal study (e.g., spaced practice), consolidates learning, and affects the development of enduring learning strategies and skills. It appears to be one of the most potent forces influencing education. Accordingly, it deserves very careful planning and considerable investment of time from educators. (p. 467).

Crooks' emphasis on the importance of the affective domain for understanding the relationship between theories of learning and educational assessment is shared by many other authors. The now established concept of 'learned helplessness' provides a good example in this respect (Petersen, Maier, & Seligman, 1993) as does McMeniman's (1989) research review of the role of motivation in relation to assessment. In the same vein, a systematic review of the research evidence concerning the impact of summative assessment on students' learning by Harlen and Deakin-Crick (2003) identified a significant relationship between self-esteem, self-efficacy, self-regulation, goal-orientation, interest, effort, locus of control and sense of self as a learner in the context of assessment. Hattie and Timperley's more recent research review on feedback (Hattie & Timperley, 2007) has provided further confirmation of the pervasive effect of assessment on a wide range of factors that affect learning. Hattie's most recent work on the relative impact of different variables of classroom life (Hattie, 2015) consistently shows that the biggest effect size out of 195 identified variables relating to influences on achievement based on a synthesis of 1200 meta-analyses, is 'teacher estimates of achievement'. This would suggest that the teacher's judgment of an individual's potential and performance has a powerful effect on learning outcomes.

Such evidence powerfully supports the argument of the focus paper that "The motivational structures that are set out by assessments need to be carefully designed, lest they motivate the wrong behaviours. Not only does washback occur under certain conditions, it is intentional and therefore it should be recognised overtly as part of the assessment design process" (p. 5).

As yet, however, there is little evidence to suggest that these kinds of theoretical insights are informing the design of assessments, suggesting perhaps that we are still in the 'pre-Enlightenment' era when it comes to assessment theory.

Are there novel theoretical or practical developments that might facilitate a better relationship between assessment and learning theory?

One of the key themes of the focal paper of this issue is that educational assessment is now a big international business. Yet, the authors suggest "learning is idiosyncratically individualistic, context-dependent and socially produced, all of which are recipes for invariance. Thus, the question that arises repeatedly is the extent to which examination scores are a product of the measuring instrument or of the underlying attributes of interest" (p. 12) leading them to speculate as to whether it is possible at all to quantify educational attainment with integrity. Are we dealing here with an emperor who has no clothes? Apparently not since, as the authors suggest, "voluminous assessment takes place annually in systematic ways in most nations" (p. 1). But if it is better than the currently conceived alternatives, it is still far from good enough. Conceptually, practically, politically, educationally, it is deeply flawed as is so well-rehearsed in the focal paper. As such the analyses presented in the paper, which embrace both the psychometric and the educational challenges confronting assessment, offer a powerful challenge to the professional assessment community.

Certainly this requires radical and creative new ways of thinking about assessment and innovative theoretical approaches that could potentially underpin a liberation from traditional concepts of quality. It should certainly involve addressing the potential of new technologies to collect performance data in significantly more comprehensive and useful ways; to make assessment more timely and personalised and designed to provide for a greater involvement of students in their assessment, since all of these characteristics have the potential to overcome some of the constraints of existing approaches and to link assessment more explicitly to educational goals.

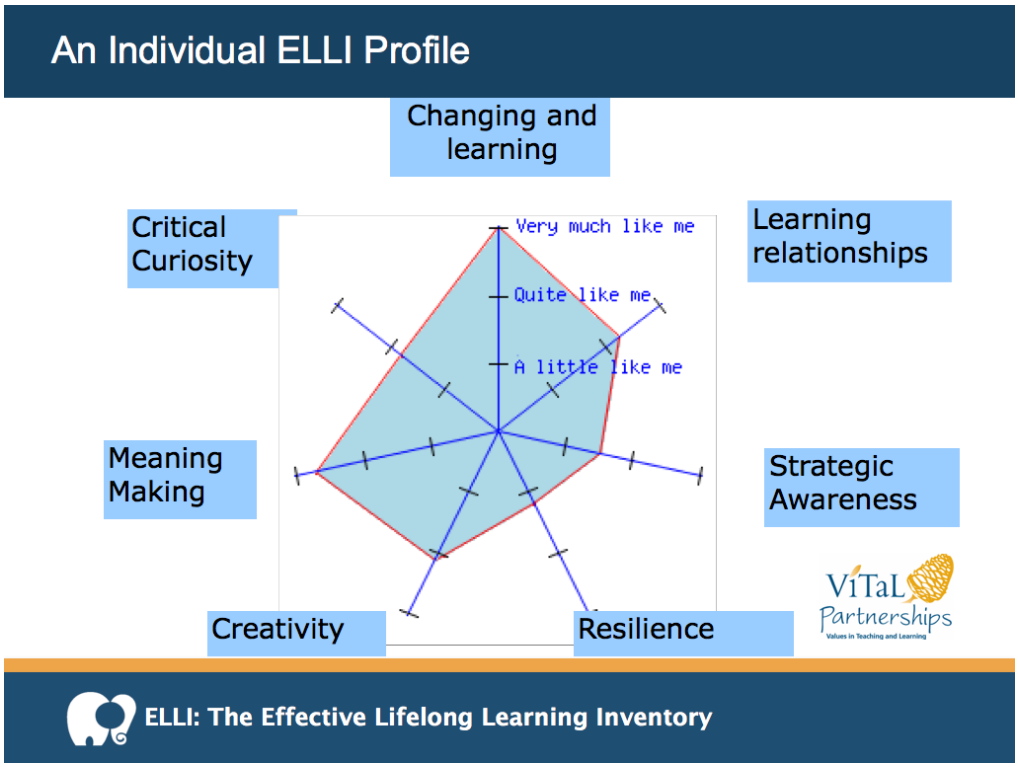
Writing in 2007, I suggested that e-assessment, like e-learning, offered the potential to provide opportunities for assessment that were quite novel (Broadfoot, 2007). Released from the considerable constraints of conventional approaches centred on paper and pencil tests of various kinds, I argued that computers offer the potential for assessment that can be more personalised and constructive and so help to build the confidence of learners, more capable of providing timely and substantive feedback, and more constructive and engaging. I suggested that such assessments can cover a wider range of learning objectives, can potentially make assessments available on demand, and provide simultaneously for both formative and summative purposes. In short, I argued that making use of computer-based technologies for assessment purposes has the potential to provide a much closer match with the research evidence about using assessment to support learning.

A more recent research review looking at the substantive progress that has been made in this area since then, makes it clear that these early arguments were valid; that technology enhanced assessment (TEA) has the potential to collect a much wider range of data on a greater variety of attributes in a timely and individualised way that can also empower students whilst saving teachers' time. (Oldfield, Broadfoot, Sutherland, & Timmis, 2014).

Moreover, just as the advent of social media and other on-line technologies has transformed the way we do business, shopping and friendship, so the gradual erosion of the traditional practical constraints governing assessment is beginning to open up new conceptual possibilities concerning how to think about assessment. Of particular relevance here is the development of the field of learning analytics, which is already showing considerable potential to provide insights into individual learning and hence to provide the possibility of targeted interventions that may be of particular value.

'Learning analytics' has been defined as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs" (Ferguson, 2012 p. 315). Its focus is on the management and use of data patterns to improve learning through the interrogation of learner-based data interaction – so-called 'key-stroke latencies' – using techniques such as predictive modelling, user profiling, adaptive learning, and social network analysis to inform decisions based on the results. Such decisions could include identifying the need for intervention, personalising support or promoting self-regulation in learning (Brown, 2012). The current collection and analysis of large, complex data sets in education is both unprecedented and becoming more and more prevalent. Moreover, learning analytics have the same potential at national and international levels. Learning analytics are already being used to provide for the more systematic analysis of aspects of the performance of a particular population or institution. (Broadfoot et al, 2013)

A particularly novel focus is that of 'disposition analytics' in which the focus is not on performance on a given task or test but rather on the characteristics of the learner. One example of the use of 'disposition analytics' is the online self-report questionnaire called ELLI (The Effective Lifelong Learning Inventory). As is shown in Figure 1 below, Elli is an on-line assessment tool that analyses and visually represents an individual's profile across seven dimensions of 'learning power' – critical curiosity, meaning making, creativity, interdependence, strategic awareness and resilience. Derived empirically through factor analysis, it is noteworthy that each of these dimensions



combines elements of 'thinking, feeling and doing' and as such provide a constructive new way of understanding the multiple influences that impact on a student's learning – cognitive certainly, but also affective and conative. This kind of approach to understanding the forces at work in learning, coupled with a flexible, accessible visual representation of each individual's profile, offers a powerful new way of linking learning and assessment, a way that takes into account much more effectively than hitherto the importance of that relationship.

Thus learning analytics can provide for assessment methods that increase agency for learners, through the use of 'learner-facing' tools that represent data back to learners and support them in the self-regulation of their learning (Siemens, 2012). Tools such as ELLI that incorporate the use of the graphical representation of a multifaceted, complex data set, also demonstrate how such data can become 'actionable' as teachers and students respond to the data presented. More generally, 'dashboard' systems, which allow students to monitor their own academic or behavioural activity

and achievements, are becoming increasingly widely used, allowing learners to access support that is relevant to them. Teachers can also use 'dashboards' to compare the progress of a particular student or class against a more generalised set of data.

These brief examples provide some hint of the unprecedented possibilities for assessment that developments in computer technology are increasingly making possible. As such they are likely to be a powerful force for change not only in assessment thinking and practice, but also in linking assessment to learning theory. Such approaches are likely to provide a stimulus for educators at all levels to think more actively about the relationship between assessment and learning. As computers gradually come to transform the way in which the curriculum is delivered – as they have transformed so many other aspects of contemporary life – so some of the more traditional sources of concern about assessment approaches such as reliability and construct validity, may well diminish in importance, to be replaced by much more holistic, objective and dynamic sources of insight about a learner's progress and performance.

But desirable as these developments may be in terms of their potential to link assessment much more closely to learning, they are not without risks. A moment's thought suggests that the potential for an even more pervasive assessment environment than that which exists at present is as dangerous as it is powerful. As Timmis, Broadfoot, Sutherland and Oldfield (2016) argue, although TEA presents unprecedented opportunities for enhancing how assessments are conducted and used, it also carries clear ethical and social risks. The starting point of the focal paper is that assessment plays a central role in education and has a huge impact on teaching and learning. Potentially more intrusive forms of assessment 'surveillance', whilst offering exciting new opportunities for capturing evidence of attainment and for providing high quality feedback, also offer a profound threat to a learner's privacy. This is particularly true all the time we continue to construct assessment devices in the politically-charged and often crude way that is common today. If the use of exponentially more powerful assessment technologies becomes widespread, the concerns about excessive intrusion into a learner's privacy that are expressed in the focal article will become significantly more important even than they are now.

It will be necessary to 'ride the tiger' of such innovation, hastening slowly in order that as new technology-enhanced assessment approaches are developed, so relevant insights into how they can be designed to enhance learning are developed in tandem. If assessment and learning are currently 'fields apart' with all the negative consequences that are the subject of this special issue, there are encouraging signs that they are likely to be drawn closer with the advent of TEA. However such

progress is unlikely to be a panacea. As the title of this response suggests, assessment can justifiably be conceived as a leviathan, a monster that holds formal education in thrall in the complex, multi-layered, often politicised business of educational assessment ably described by Baird et al.

New technologies offer the potential for a much closer link between this leviathan and educational purposes, between the theories and practices of assessment and the business of teaching and learning. The question, however, remains as to whether this will result in the taming of the leviathan or rather, whether it might rather increase its stranglehold. If individuals find themselves caught up yet more hopelessly in the coils of a system they can do nothing to resist or control, such change is a terrifying project. Assessment scholars will require the utmost vigilance if future gains are not to be outweighed by the new problems created. This special issue is therefore a most welcome call to arms that the long-neglected silences concerning the crucial relationship between assessment and learning theory needs urgently to be addressed. To the extent that the paper challenges some of the basic assumptions of psychometric approaches to measurement, that it highlights the yawning omissions in how assessment data are used, and the pressing need for more than pragmatic convictions about the value of AfL, the paper challenges all those interested in making assessment activities work better to support learning.

Hopefully, for all the reasons set out in this response, this is the dawn of an exciting new era for educational assessment; an era in which, with care, the assessment leviathan that currently constrains and shapes virtually all formal educational activity may finally be tamed, harnessed in the service of learning, rather than preying upon it.

References

Broadfoot, P. (2007). *An introduction to assessment*. London, UK: Continuum.

Broadfoot, P., Timmis, S., Payton, S., Oldfield, A., & Sutherland, R. (2016). *Rethinking assessment in a digital age: Opportunities, challenges and risks*. Bristol, UK: University of Bristol Graduate School of Education.

Brown, M. (2012). *Learning analytics: Moving from concept to practice*. Denver, CO: Educause Learning Initiative.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58*(4), 438-481.

Dweck, C. S. (1989). Motivation. In A. Lesgold & R. Glaser (Eds.), *Foundations for a psychology*

of education (pp. 87-136). HIllsdale, NJ: Lawrence Erlbaum Associates.

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning, 4*(5/6), 304-317.

Harlen, W., & Deakin Crick, R. (2003). Testing and motivation for learning. *Assessment in Education: Principles, Policy & Practice, 10*(2), 169-207.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112.

Hattie, J. (2015). The applicability of Visible Learning to higher education. *Scholarship of Teaching and Learning in Psychology, 1*(1), 79-91.

McMeniman, M. (1989). Motivation to learn. In P. Langford (Ed.), *Educational psychology – An Australian perspective* (pp. 215-239). Melbourne, Australia: Longman Cheshire.

Oldfield, A., Broadfoot, P., Sutherland, R., & Timmis, S. (2016). *Assessment in a digital age: A research review*. Bristol, UK: University of Bristol Graduate School of Education.

Peterson, C., Maier, S. F., & Seligman, M. E. P. (1993). *Learned helplessness: A theory for the age of personal control*. New York, NY: Oxford University Press.

Siemens, G. (2012, April 29 - May 2). *Learning analytics: Envisioning a research discipline and a domain of practice*. Paper presented at the Second International Conference on Learning Analytics & Knowledge, Vancouver, BC.

Timmis, S., Broadfoot, P., Sutherland, R., & Oldfield, A. (2016). Rethinking assessment in a digital age: Opportunities, challenges and risks. *British Educational Research Journal, 42*(3), 454-476.